

# Polysemy in Lexical Semantics -- Automatic Discovery of Polysemous Senses and Their Regularities

Noriko Tomuro and Steve Lytinen  
DePaul University, College of Digital Media  
243 S. Wabash Ave.  
Chicago, IL 60604 U.S.A.  
{tomuro,lytinen}@cs.depaul.edu

Lexical ambiguity poses a significant problem in Natural Language Processing (NLP). In processing texts, polysemous words may hinder the accuracy of the derived results, because different contexts are mixed in the instances, collected from the corpus, in which a polysemous word occurred. At DePaul University, we have been working on techniques to automatically discover the individual senses of polysemous words and the regularities that exist between the senses. Then our ultimate goal is to develop a lexical representation which encodes various relations between word senses, and build a lexicon which can support knowledge-intensive, semantic NLP tasks such as textual entailment and question-answering.

## Automatic Discovery of Polysemous Word Senses

In the NLP research, many techniques have been developed which discover word senses or semantic classes from corpora (e.g. [1]). Most of them use distributional similarity to cluster words, where members in the same cluster are considered to be synonyms. However, such techniques do not apply to polysemous words directly, because words which have a similar polysemous pattern end up in the same cluster, thus the individual senses cannot be discovered. To remedy the problem, we have developed a new clustering algorithm which can discover senses of polysemous (as well as monosemous) words [2]. Our algorithm is novel in that it incorporates the similarity between features (rather than assuming independence) in clustering instances. We applied the algorithm to English and Japanese un-annotated corpora, and discovered polysemous adjectival senses. Our next task is to refine the algorithm (in particular the parameters), and apply it to other parts of speech for further evaluation.

## Automatic Discovery of Systematic Polysemy

When a word has multiple senses, some senses may be un-related (e.g. “bank” as a financial institution or sloping land), while others may be related (e.g. “chicken” as an animal or meat of the animal). Some of those relations are regular, in that the same relation is observed in several other polysemous words (a linguistic phenomenon called *systematic polysemy* or *regular polysemy* [3]). Also, some relations exist simultaneously in a discourse (e.g. “increase” as a quantity and/or a process). To discover systematic polysemy automatically, we have developed a technique which groups related word senses that are organized in a hierarchy (in particular, WordNet nouns) [4]. Our next project is to discover systematic polysemy from corpus (rather than an existing lexicon), by first applying the aforementioned word sense discovery algorithm to an un-annotated corpus, then developing an automatic technique that identifies various systematic relations between the senses.

[1] P. Pantel and D. Lin. Discovering Word Senses from Text. In *Proceedings of the 8th ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2002.

[2] N. Tomuro, S. Lytinen, K. Kanzaki and H. Isahara. Clustering Using Feature Domain Similarity to Discover Word Senses for Adjectives. In *Proceedings of the 1st IEEE International Conference on Semantic Computing (ICSC)*, 2007.

[3] J. Pustejovsky. *The Generative Lexicon*. The MIT Press, 1995.

[4] N. Tomuro. Tree-cut and A Lexicon based on Systematic Polysemy. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2001.